

ARTÍCULO ORIGINAL



Uso del lenguaje *R* para el procesamiento de datos y la generación de mapas sobre COVID-19

*Use of the R Language for Data Processing
and the Generation of Maps about COVID-19*



Alvaro Reinoso Lorente

arlorente@uclv.cu • <http://orcid.org/0000-0002-3064-1701>

DELEGACIÓN DEL MINISTERIO DE CIENCIA, TECNOLOGÍA Y MEDIO AMBIENTE EN VILLA CLARA, CUBA

Romel Vázquez Rodríguez

romel@uclv.edu.cu • <https://orcid.org/0000-0003-3313-2459>

Carlos Pérez Risquet

cperez@uclv.edu.cu • <https://orcid.org/0000-0002-3547-3049>

UNIVERSIDAD CENTRAL "MARTA ABREU" DE LAS VILLAS, CUBA

Recibido: 2020-09-06 • *Aceptado: 2020-10-06*

RESUMEN

En la actualidad el mundo se encuentra sumergido en la pandemia global "COVID-19". El estudio del comportamiento de esta enfermedad genera mucha información, que en la mayoría de los casos puede ser georreferenciada, con el objetivo de realizar un análisis espacio-temporal para arribar a conclusiones. Tradicionalmente se han empleado los Sistemas de Información Geográfica para procesar y analizar información georreferenciada; sin embargo, actualmente existen otras alternativas para este fin, como es el caso del lenguaje *R*, que se ha convertido sin duda alguna en una importante herramienta para el análisis espacio-temporal y generación de mapas, debido a que cuenta con una gran cantidad de técnicas estadísticas que facilitan el procesamiento de datos y la generación de mapas. La evolución de esta enfermedad en cada región o país, es sin lugar a dudas de interés para muchas investigaciones. Analizar cómo se desarrolla la enfermedad en el transcurso de días, semanas, meses y años es de gran importancia. En este trabajo se muestran algunos ejemplos de cómo se puede utilizar el lenguaje *R* para la generación y animación de mapas, que permiten analizar la evolución de una enfermedad como la COVID-19, así como otros datos relacionados.



PALABRAS CLAVE: COVID-19; lenguaje R; mapas de coropletas; mapas de símbolos graduados.

ABSTRACT

Currently the world is submerged in the global pandemic “COVID-19”. The study of the behavior of this disease generates a lot of information that in most cases can be georeferenced in order to carry out a spatio-temporal analysis to reach conclusions. Geographic Information Systems have been traditionally used to process and analyze georeferenced information. However, currently there are other alternatives for this purpose such as the R language, which undoubtedly became an important tool for spatio-temporal analysis and generation of maps, because of its large number of statistical techniques that facilitate data processing and map generation. The evolution of this disease in each region or country is certainly of interest to many investigations. Analyzing how the disease develops over the course of days, weeks, months and years is of great importance. This work shows some examples of how the R language can be used for the generation and animation of maps that allow analyzing the evolution of a disease such as COVID-19 as well as other related data.

KEYWORDS: *choropleth maps; COVID-19; graduated symbol maps; R language.*

INTRODUCCIÓN

La creciente ola de información que se genera diariamente es almacenada en diversos formatos de datos. Muchos de estos formatos han sido creados por instituciones o comunidades, en dependencia de la necesidad de almacenamiento que estos poseen para sus grandes volúmenes de datos. Gran parte de esta información almacenada está relacionada con variables climáticas y datos asociados a otros eventos que ocurren cotidianamente, con el objetivo de analizarlos y poder arribar a conclusiones sobre el tema.

La mayoría de estos datos están relacionados con el medio ambiente, enfermedades, entre otros. Muchos poseen componente espacial, lo que significa que estos datos pueden ser georeferenciados, o lo que es lo mismo, tienen una localización geográfica (Vázquez Rodríguez, 2015).

Gran parte de los elementos que existen en la naturaleza pueden ser representados mediante formas geométricas (puntos, líneas o polígonos), lo que suele representarse en el mundo de los Sistemas de Información Geográfica (SIG) mediante formatos de datos de tipo vectorial. Comúnmente se selecciona una estructura de datos vectorial cuando hay que reflejar

entidades reales ubicadas en el espacio (carreteras, ríos, regiones), también cuando se quiere realizar un análisis entre unidades espaciales donde se manejan datos sobre una extensión de territorio que se considera homogénea (Barroso & Puebla, 1997; Murphy, 1995).

En las últimas décadas se ha producido un desarrollo importante de la Estadística espacio-temporal, dada la necesidad de analizar la evolución temporal del comportamiento espacial de magnitudes aleatorias que son de interés en estudios desarrollados en diversas áreas aplicadas, tales como Medioambiente, Geofísica, Biología y Medicina (Hernández-Leal, Duque-Méndez, & Moreno-Cadavid, 2017). En particular, en el contexto geo-estadístico, se ha producido un avance importante en la derivación de modelos flexibles para el procesamiento y análisis estadístico de datos espacio-temporales. Los enfoques adoptados se fundamentan esencialmente en la estadística espacial, dado su auge previo en este contexto (Salmerón Gómez, *et al.*, 2008). R como lenguaje estadístico (Lamigueiro, 2014) cuenta con diversos paquetes para manipulación de datos vectoriales y datos espacio-temporales. Estos se encuentran organizados en vistas o temas en un repositorio oficial conocido como *The Comprehensive R Archive Network* (CRAN). El análisis espacial es una extensión natural a las capacidades de R, que es primordialmente un lenguaje de programación para análisis de datos, por lo que cuenta con un sin fin de técnicas estadísticas que no se encuentran tradicionalmente en plataformas SIG.

El Coronavirus 2019 (COVID-19) es una enfermedad infecciosa causada por el virus del Síndrome Respiratorio Agudo Severo tipo-2 (SARS-CoV-2), que fue declarado pandemia en marzo de 2020. Las tasas de letalidad se estiman entre el 1% y el 3% de los infectados por el virus, afectando principalmente a los adultos mayores y a aquellos con comorbilidades, como hipertensión, diabetes, enfermedad cardiovascular y cáncer. Diariamente se generan altos volúmenes de datos relacionados con esta pandemia, debido a la alta cifra de contagios que presenta y el interés mundial que ha despertado. El estudio del comportamiento esta enfermedad en un espacio y período dado es de vital importancia para su enfrentamiento (Díaz-Castrillón & Toro-Montoya, 2020).

El objetivo de este trabajo es procesar datos de la enfermedad COVID-19 y visualizarlos en forma de mapas de coropletas y mapas de símbolos graduados, de forma tal que posibilite un análisis espacial y temporal de los datos, mediante secuencias de imágenes. Se utiliza el lenguaje R para el procesamiento de datos, la generación y animación de mapas. El procedimiento descrito puede adaptarse sin mucha complejidad para ser utilizado con otros datos en diferentes regiones y con otras escalas temporales.

METODOLOGÍA

Para la generación de los mapas, se parte de datos disponibles, previamente elaborados con otros fines, algunos de los cuales están disponibles en Internet y otros fueron solicitados a diferentes entidades dentro de las cuales se encuentra el Ministerio de Salud Pública de la República de Cuba (Minsap). En todos los casos es necesario procesar dichos datos de acuerdo al

tipo de análisis que se desee realizar, para luego asociarlos a las geometrías correspondientes en el formato *shapefile* (shp), un tipo de modelo vectorial. Con dicha asociación ya estamos en condiciones de generar los diferentes mapas temáticos (símbolos graduados y coropletas). En este caso se utiliza el lenguaje R para realizar el procesamiento de los datos y la generación de los mapas.

Seguidamente, en esta metodología se describe el modelo vectorial como dato espacial y se menciona el formato *shapefile*, a partir del cual se generan los mapas. Por último, se describe el lenguaje R como herramienta utilizada para el procesamiento y generación de mapas temáticos.

Los datos espaciales permiten una representación del mundo real o cotidiano. Dicha representación trae consigo la incorporación de modelos dinámicos, dependientes de las variables temporales y espaciales. Una gran cantidad de estudios han permitido acrecentar las capacidades y condiciones de desarrollo para herramientas que realicen análisis de variables temporales. Para modelar y entender la evolución de los fenómenos geográficos el tiempo es determinante, se considera que las transformaciones del espacio geográfico dependen de las actividades y los fenómenos espacio-temporales (Durango Vanegas, 2013).

MODELO VECTORIAL

El modelo vectorial permite modelar el espacio geográfico mediante puntos, líneas o polígonos, si se asocian a estas geometrías una serie de valores definitorios. Al definir las formas geométricas básicas, todas ellas pueden, en última instancia, reducirse a puntos. Así, las líneas son un conjunto de puntos interconectados en un determinado orden, y los polígonos son líneas cerradas, también expresables como una serie de puntos. Todo elemento del espacio geográfico queda definido por una serie de puntos que determinan sus propiedades espaciales y una serie de valores asociados. Una única entidad puede contener varias primitivas, esto permite representar territorios en los que no se puede mostrar toda su extensión mediante una sola primitiva. Una entidad que esté compuesta por más de una primitiva, contiene un único conjunto de valores asociados a ella (Olaya, 2014).

El modelo de representación vectorial posee como elemento particular la topología. De manera informal, la topología se ocupa de aquellas propiedades y características de las figuras o cuerpos geométricos que permanecen invariantes cuando dichas figuras son plegadas, dilatadas, contraídas o deformadas, de manera que no aparezcan nuevos puntos o se hagan coincidir puntos diferentes (Stadler, 2002).

La representación de puntos o líneas es inmediata, sin embargo, al representar polígonos aparecen dos situaciones diferentes (Stadler, 2002):

- Si los polígonos aparecen aislados los unos de los otros, cada polígono se codifica como una línea cerrada, esta representación se denomina modelo *Orientado a objetos*. Esta representación tiene como desventaja que, si los polígonos se yuxtaponen, codificar los polígonos como líneas cerradas tiene el problema de que habría que repetir cada una de las líneas interiores.

- El formato alternativo es el modelo *Arco-Nodo*, donde se codifican las líneas por separado, y posteriormente se define cada uno de los polígonos a partir del conjunto de líneas que lo componen.

Generalmente se considera al formato vectorial más adecuado para la representación de entidades o variables cualitativas; sin embargo, esto no siempre es necesariamente así (Stadler, 2002).

Para representar variables cualitativas y objetos en el modelo vectorial existen dos formas posibles (Stadler, 2002):

- Formato vectorial *Arco-Nodo*.
- Formato vectorial *Orientado a objetos*, menos adecuado pues se introduce mucha información redundante.

FORMATO SHAPEFILE

El formato *shapefile* es un formato de archivos para datos vectoriales desarrollado por la compañía ESRI y es ampliamente utilizado. A diferencia de otros formatos de datos existentes, está compuesto por al menos por tres tipos de ficheros, los cuales tienen las extensiones siguientes (Richards, *et al.*, 2010):

- *.shp*: almacena las entidades geométricas de los objetos.
- *.shx*: almacena el índice de las entidades geométricas.
- *.dbf*: almacena la tabla de atributos de los objetos.

Es común encontrar también un cuarto archivo con extensión *.prj*, que contiene la información referida al sistema de coordenadas (Mas, 2018).

Este formato permite almacenar solamente una geometría por tabla, no puede almacenar valores nulos, redondea los números y en él no se pueden almacenar fecha y hora en un mismo campo.

Actualmente no solo los SIG son capaces de manipular satisfactoriamente archivos *shapefile*; han surgido lenguajes muy efectivos para su manipulación, como es el caso del lenguaje R (Richards, *et al.*, 2010).

Las imágenes resultantes de la generación de mapas a partir de un archivo *shapefile* contienen mapas temáticos, que por lo general son de coropletas o de símbolos graduados, ya que estos tipos de mapas facilitan la comparación de una medida estadística de una región con otra, o muestra la variabilidad de esta para una región dada.

MAPAS TEMÁTICOS

Los mapas temáticos son mapas basados en mapas topográficos, cuyo fin es representar cualquier fenómeno geográfico de la superficie terrestre, tanto cualitativo como cuantitativo. Un mapa temático es en buena medida el final de un proceso investigativo. De esta forma el mapa será la suma de las fuentes y de la propia aportación o interpretación personal de un investigador (Bravo, 2000).

Estos tipos de mapas son una herramienta eficaz utilizada por los especialistas para realizar análisis comparativos de tendencias y variedades presentes en la información de una

determinada fuente de datos, debido a que tratan los valores en un campo designado estadísticamente y asignan los estilos a las clases identificadas según los parámetros que el usuario especifica. Se pueden representar datos derivados a la densidad poblacional, tasas de mortalidad, enfermedades, etc., y todo esto codificado bien sea por colores o gráficas (Bravo, 2000).

MAPAS DE SÍMBOLOS PROPORCIONALES O GRADUADOS

Los mapas de símbolos graduados son un tipo de mapas de puntos, donde la presencia de símbolos en determinadas posiciones, está relacionada con un tamaño que es proporcional a la magnitud de la cantidad del atributo que se pretende visualizar. De esta forma es fácil establecer comparaciones visuales entre varios valores de un atributo único representado en el mapa (Taylor & Brower, 1984). El mapa B en la figura 1 se muestra un ejemplo de un mapa de símbolos graduados.

Este tipo de mapa puede ser creado utilizando una regla que contenga un tipo de geometría puntual para definir un símbolo, por ejemplo, cuadrado o círculo, donde su tamaño puede variar de acuerdo a la proporción de la cantidad del valor que se muestra. De esta forma, los símbolos tendrán tamaños diferentes acordes al valor del atributo donde serán representados (Taylor & Brower, 1984).

Es muy común que este tipo de mapa sea utilizado sobre todo para la representación de datos socioeconómicos, pero en la práctica mediante este sistema se puede representar cualquier dato, incluyendo totales, proporciones y razones (Taylor & Brower, 1984). Además del tamaño, en un símbolo se pueden emplear también otras metáforas visuales como forma, color, luminosidad, transparencia, textura, etc., para codificar otros atributos numéricos asociados a diferentes variables observadas en esa misma localización geográfica.

Mapas de coropletas

Los mapas de coropletas son una variedad de mapas temáticos, donde un área geográfica es dividida en diferentes regiones, cada una de las cuales se distingue de las restantes a través de un color correspondiente a un determinado atributo. De esta manera proporcionan una forma fácil de visualizar una variación de un único atributo de una región determinada. En la figura 1 se muestra un ejemplo de este tipo de mapa temático. Normalmente, se crean intervalos dentro de una gama de valores que un atributo puede tomar, y para cada uno de estos intervalos se le atribuye un color o tonalidades de colores diferentes. Para la creación de este tipo de mapas se puede definir una regla con un filtro por clase, que será caracterizada por un elemento o región que contenga un tipo de geometría poligonal. Cada color representa una determinada clase de valores (Sherman, Gartin, & Burger, 1989).

Cuando se crean estos tipos de mapas, el usuario debe seleccionar el tipo de rango a utilizar para representar la distribución del hecho, por ejemplo, intervalos iguales, conteos iguales (cuantiles), rupturas naturales, desviación estándar o rango personalizado. Esta libertad para escoger un rango de cualquier tipo origina una variabilidad en cualquier mapa creado para identificar determinados hechos (Harries, 1995).

Un problema adicional con los mapas de áreas definidas por límites geográficos es el relacionado con el “Problema de la Unidad de Área Modificable”, en el cual los cambios en las áreas geográficas definidas por límites empleadas para representar temáticamente la distribución de un atributo, pueden afectar e incluso ocasionar equivocaciones en la interpretación del mapa. Sin embargo, el método de mapeo temático por áreas definidas por límites no debe ser completamente desechado. Los mapas temáticos de áreas definidas por límites geográficos son importantes, ya que las áreas que representan frecuentemente son regiones geográficas usadas para propósitos políticos y administrativos (Bailey & Gatrell, 1995).

EL LENGUAJE R PARA LA VISUALIZACIÓN DE ARCHIVOS SHAPEFILE

El lenguaje *R* es una plataforma de análisis estadístico. Es un referente desde hace varios años en este campo, debido a que cuenta con herramientas gráficas muy avanzadas. Se puede obtener gratuitamente, ya que es parte del proyecto colaborativo de *software* libre GNU y se distribuye bajo la Licencia Pública General. Esta licencia tiene por propósito declarar que el *software* es libre y protegerlo de intentos de apropiación que restrinjan esas libertades a nuevos usuarios, cuando el *software* es distribuido o modificado. Por ser un programa de código abierto, *R* es fruto del esfuerzo de miles de personas en todo el mundo que colaboran en su desarrollo, lo que permite solucionar problemas de errores de programación rápidamente, así como el desarrollo de paquetes modulares, que son complementos especializados para temas específicos. Estos paquetes se basan a menudo en métodos muy innovadores, permitiendo utilizar *R* para una amplia gama de problemas. Existen paquetes para procesar datos tan diversos como censos, series de tiempo, secuencias genéticas o información económica, así como implementar una gran variedad de métodos estadísticos. Durante los últimos años, se han creado diferentes paquetes dirigidos al análisis espacial, convirtiendo al lenguaje *R* en una potente herramienta para llevar a cabo el mapeo y el análisis de todo tipo de información georeferenciada (Mas, 2018).

Aspectos generales del lenguaje R

El lenguaje *R* se caracteriza como un sistema completamente diseñado y vinculado directamente al análisis de datos. Generalmente se utiliza *R* como un sistema estadístico, sin embargo, la descripción correcta es la de un lenguaje en el que se han implementado muchas funciones con fines estadísticos. Algunas están incluidas en el entorno base de *R* y otras se encuentran en forma de bibliotecas (Bellosta, 2018; Mas, 2018; Paradis & Ahumada, 2003; Santana & Farfán, 2014).

Actualmente se encuentran disponibles alrededor de 14.000 bibliotecas desarrolladas en *R*, que cubren diversos campos, desde aplicaciones Bayesianas, financieras, generación de mapas, *wavelets*, análisis de datos espaciales, etc. (Bellosta, 2018; Mas, 2018; Paradis & Ahumada, 2003; Santana & Farfán, 2014).

El lenguaje R como una plataforma SIG

Desde ya varios años se han desarrollado paquetes en *R* dirigidos al análisis espacial, permitiendo la manipulación y el análisis de datos espacio-temporales en este lenguaje (Mas, 2018).

Entre las ventajas del lenguaje *R* para la realización de visualizaciones y análisis espaciales se encuentran:

- *R* permite repetir y replicar todos los procesos una vez que se han desarrollado, lo que genera de manera automática una investigación transparente.
- *R* es primordialmente un lenguaje de programación para análisis de datos, por lo que cuenta con un sin fin de técnicas estadísticas, que no se encuentran tradicionalmente en plataformas SIG (Mas, 2018).

La mayor desventaja de usar *R* por sí solo como SIG es su capacidad para la visualización de datos espaciales, que por lo general solo permite la generación de imágenes estáticas. Si se desea explorar un mapa de manera dinámica, se deben utilizar funcionalidades específicas (Mas, 2018). Por esto es recomendable parear *R* con un *software* SIG especializado, por ejemplo QGIS, que es sumamente poderoso y de código abierto (Mas, 2018).

Paquetes para la manipulación de formatos científicos de datos espacio-temporales y geoespaciales en R

R forma parte de un proyecto colaborativo y abierto, que promueve la creación de paquetes con el fin de brindar la posibilidad de compartir el trabajo y contribuir además al crecimiento del propio lenguaje. Existe un repositorio oficial de paquetes, que son organizados en vistas (o temas) según su naturaleza y función. El lenguaje cuenta con paquetes para facilitar la manipulación de datos científicos y espacio-temporales como son: *Raster* (Hijmans, 2013), *rgdal* (Keitt, Rowlingson, & Bivand, 2013), *tidyr* (Messy, 2019), *ClassInt* (Bivand, 2013), *Maptools* (Bivand, *et al.*, 2016), *RcolorBrewer* (Neuwirth, 2013), *ggplot2* (Wickham & Chang, 2015), *Plyr* (Wickham, 2020).

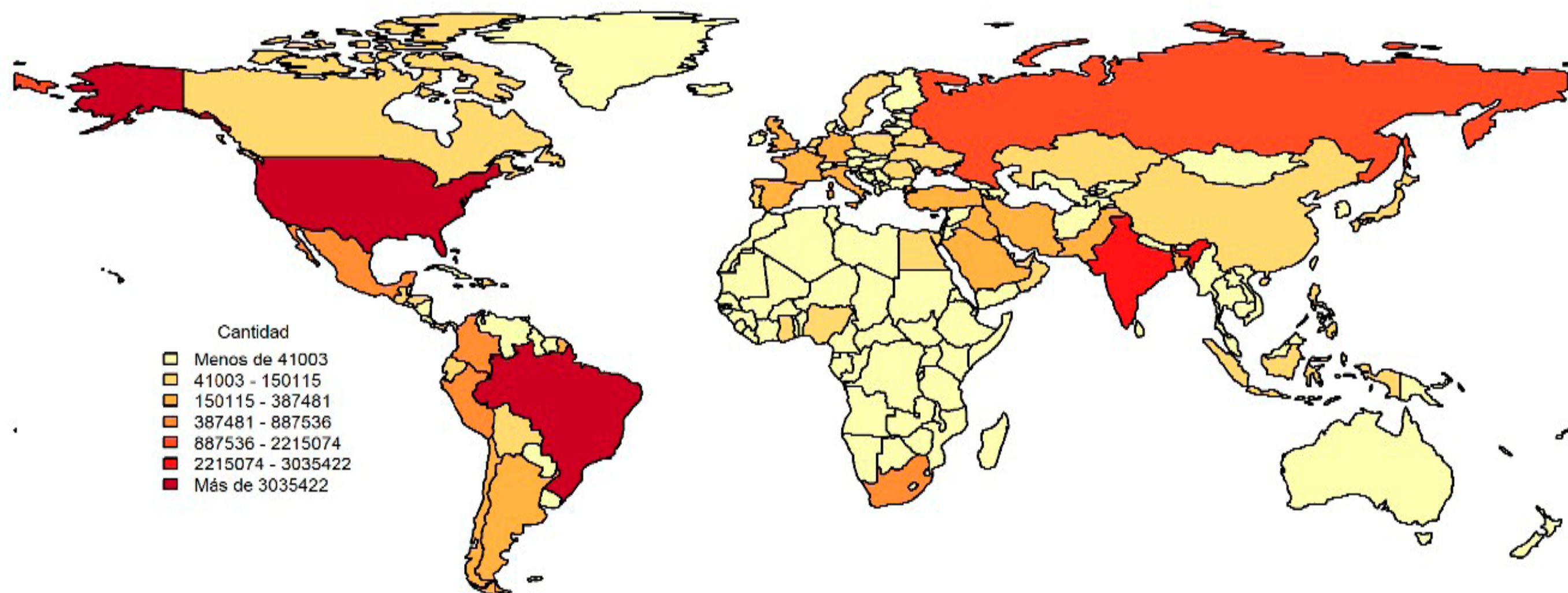
RESULTADOS Y DISCUSIÓN

En esta sección se describen los principales resultados obtenidos sobre la generación de mapas para la COVID-19 en el lenguaje *R*: generación de mapas de coropletas mundiales con los casos positivos y fallecidos por países a causa de la pandemia; procesamiento de datos de casos positivos de Cuba a nivel provincial y municipal, los cuales se visualizan en forma de mapas de coropletas y de símbolos graduados, incluyendo animaciones temporales a partir de los mapas de símbolos graduados; visualización en mapas de los datos de la autopesquisa, en conjunto con otras instituciones y entidades del país como el Minsap, la empresa Geocuba y el Ministerio de Ciencia, Tecnología y Medio Ambiente de la República de Cuba (Citma).

GENERACIÓN DE MAPAS DE COROPLETAS A NIVEL MUNDIAL CON LOS CASOS POSITIVOS Y FALLECIDOS POR PAÍSES A CAUSA DE LA COVID-19

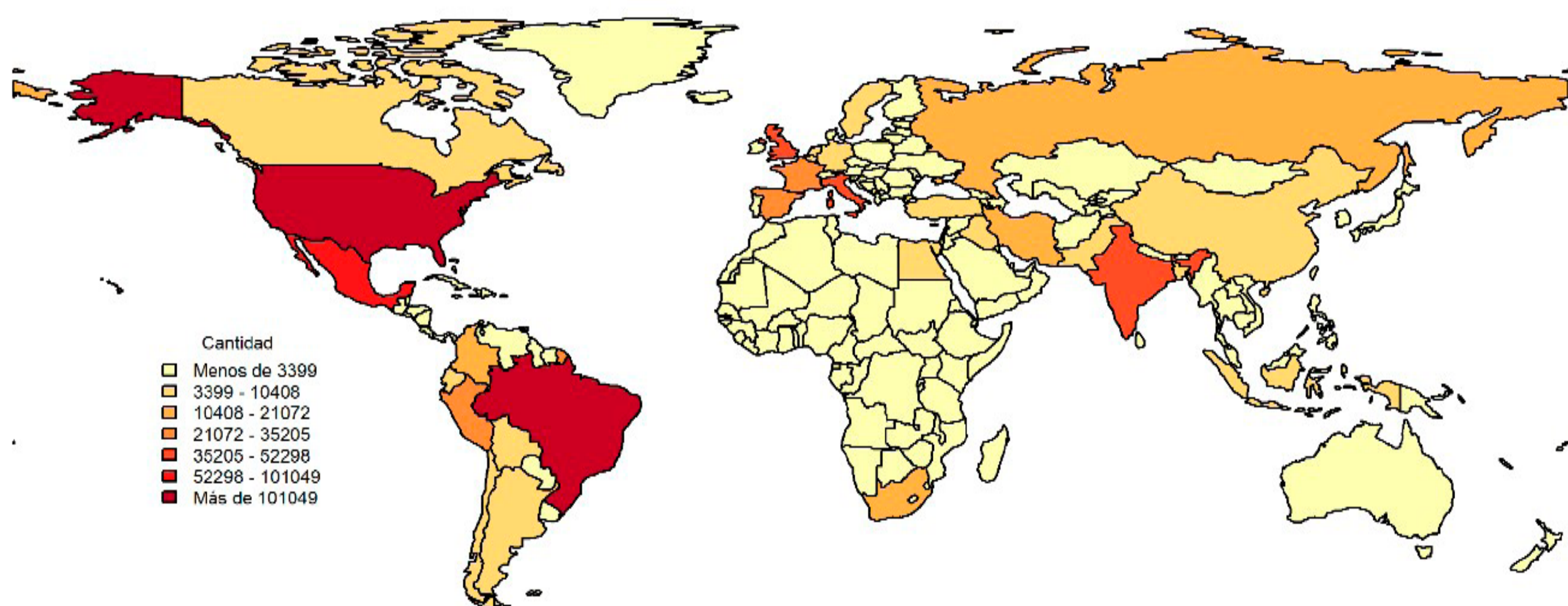
El primer resultado que se obtuvo en la visualización de mapas sobre la COVID-19 está relacionado con la generación de mapas a nivel mundial de datos oficiales de la Organización Mundial de la Salud.

Cantidad de Casos a nivel mundial



Mapa A

Cantidad de Fallecidos a nivel mundial



Mapa B

Figura 1. Mapas del Mundo con casos positivos y fallecidos de la COVID-19.

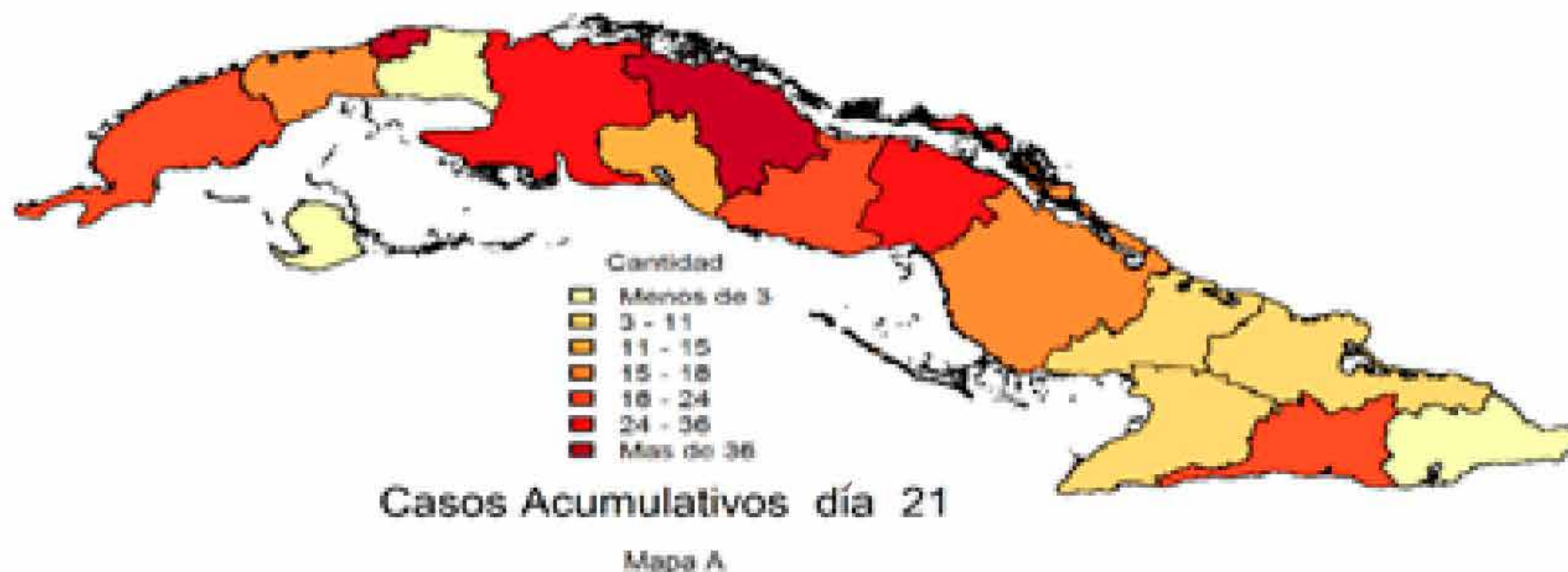
A partir de un pequeño fragmento de código en *R*, disponible en el sitio Web del Centro Europeo para la Prevención y el Control de Enfermedades (<https://www.ecdc.europa.eu>) se logra obtener un documento de *Microsoft Excel* con los datos oficiales de casos y fallecidos de COVID-19 por países. Tomando como base el fragmento de código mencionado anteriormente se desarrolló un *script*, el cual utiliza el documento de *Excel* obtenido y un fichero *shapefile* con la geometría de los países a nivel mundial para generar los mapas con las cantidades de casos positivos (ver figura 1 *Mapa A*) y de fallecidos (ver figura 1 *Mapa B*). En este caso los mapas de la figura fueron generados calculando siete clases de intervalos a partir del método *Jenks* (Intervalos Naturales) y con una paleta de colores de amarillo (menor cantidad de casos) a rojo (mayor cantidad de casos). No obstante, se pueden utilizar otros métodos, otro número de clases, así como otros tipos de paletas de colores.

GENERACIÓN DE MAPAS DE COROPLETAS Y DE SÍMBOLOS GRADUADOS DE CASOS POSITIVOS A LA COVID-19 EN CUBA

En el caso de la generación de mapas de COVID-19 en Cuba se utilizó como base la división político administrativa a nivel de provincias, y los datos oficiales publicados por el Ministerio de Salud Pública de la nación. Se generaron mapas de coropletas y de símbolos graduados con las cantidades acumuladas de casos por provincias por cada día desde el inicio de la epidemia, con el objetivo de analizar su desarrollo en el tiempo.

En la figura 2 *Mapa A* se muestra uno de los mapas de coropletas generados, correspondiente al día 21 de afectación. Estos tipos de mapas se generan mediante un *script* similar a los encargados de generar los mapas explicados en la sección anterior. A este *script* se le agrega un fragmento de código, que se encarga de procesar los datos de casos positivos, con el objetivo de agruparlos por provincias y por día de reporte, para así visualizarlos temporalmente.

Cantidad de Casos en Cuba a nivel Provincial



Mapa de los casos confirmados de Covid 19 en Cuba por Provincia

Casos acumulativos día 21

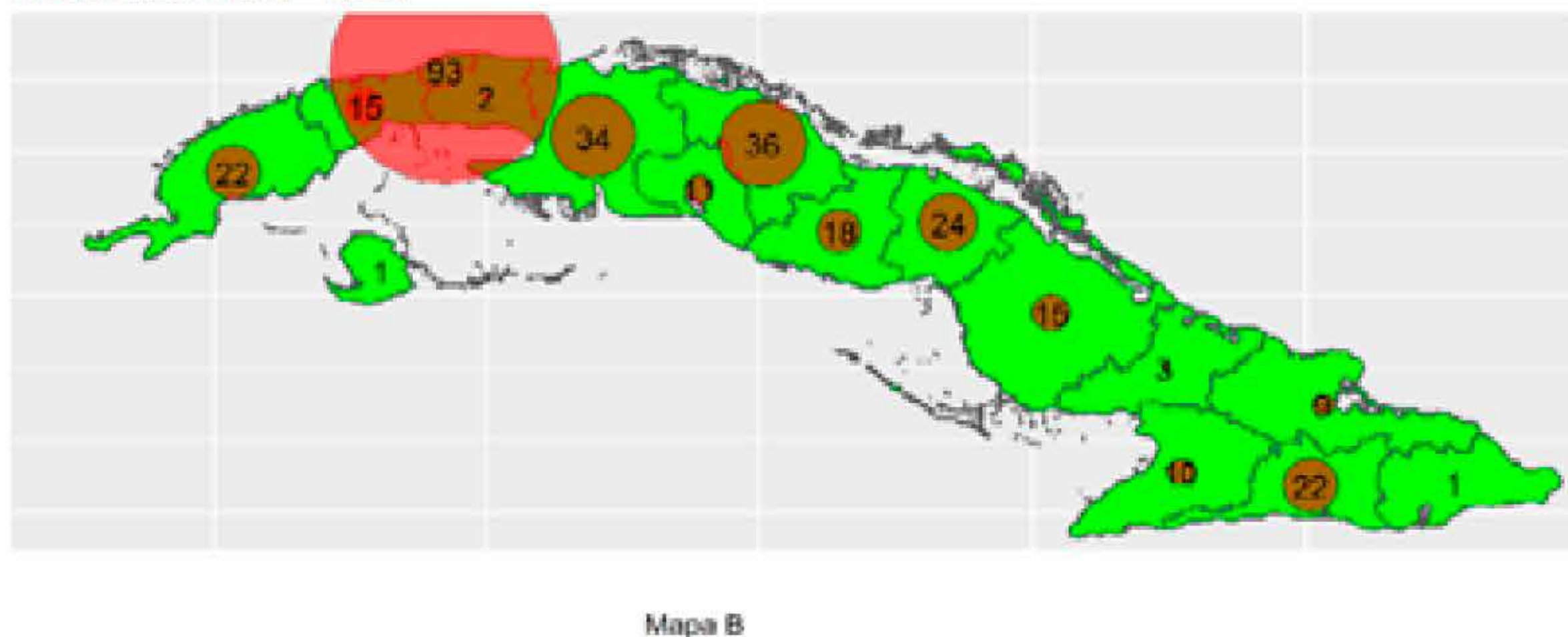


Figura 2. Mapa de Cuba con casos positivos a la COVID-19.

En la figura 2 *Mapa B* se muestra una imagen del mapa de símbolos graduados correspondiente al acumulado de casos positivos por provincia hasta el día 21 de la epidemia. Lo más importante en este caso es que con la secuencia de mapas de símbolos graduados se puede hacer una animación en forma de archivo *CompuServe Gif* (.gif) animado, donde se puede

comprobar el incremento de los casos al pasar los días. Para generar estos mapas se creó un *script*, que a partir de los datos de casos positivos genera una tabla con los casos agrupados por provincias y por día de reporte. Luego acopla la tabla a un fichero *shapefile* con la geometría de las provincias de Cuba y obtiene un centroide, el cual se utiliza para calcular el punto medio de las provincias, donde sitúa los símbolos graduados de tipo círculo, cuyo diámetro se corresponde con la cantidad de casos de esa provincia en ese día. El fichero *.shp* se enlaza con la ubicación de los símbolos graduados y comienza la generación de mapas por cada día de reporte. Una vez generados los mapas, se crea un archivo *.gif* animado con todos ellos, que acopla todas las imágenes ordenadas por fecha, de manera tal que al ejecutarse el *.gif* se visualizan los mapas desde el primer día en que se reportaron casos hasta el último, mostrando la evolución de la enfermedad a lo largo del tiempo. De esta manera los especialistas pueden realizar un estudio del comportamiento de la enfermedad en las diferentes regiones.

De manera similar al caso de generación mapas de coropletas de Cuba, se desarrolló un *script*, el cual visualiza los datos de casos y de fallecidos de la COVID-19 para la provincia de Villa Clara por municipios. El *script* desarrollado también incluye el procesamiento de datos y es perfectamente adaptable para el resto de las provincias.

GENERACIÓN DE MAPAS DE COROPLETAS POR MUNICIPIOS Y ÁREAS DE SALUD A PARTIR DE LOS DATOS DE LA AUTOPESQUISA EN VILLA CLARA

Como parte de la colaboración de los autores de este trabajo con el Consejo de Defensa Provincial de Villa Clara y de la ayuda que brindó la Universidad Central “Marta Abreu” de Las Villas para el procesamiento de datos generados por la aplicación móvil *Pesquisador Virtual* (<https://www.apklis.cu/application/cu.online.survey/>), puesta a disposición de la población por parte de la Universidad de las Ciencias Informáticas en coordinación con el Ministerio de Salud Pública y el Ministerio de Comunicaciones de la República de Cuba, se pudo obtener de la Dirección de Provincial de Salud de Villa Clara los datos recopilados de las personas que se autopesquisaron. Estos datos fueron de gran importancia para la provincia, y en particular para las áreas de salud que atiende cada uno los policlínicos más importantes. En la aplicación de autopesquisa toda persona debe especificar el policlínico a que pertenece y existe una relación biunívoca entre el policlínico al que pertenece cada persona y el área de salud. Esta información fue utilizada para hacer una correspondencia y generar mapas en formato *shapefile* con la información recopilada de las autopesquisas a nivel de áreas de salud.

El procesamiento de los datos de las autopesquisas fue realizado en *R* y la generación de los *shapefile* fue realizada también en este lenguaje.

Posteriormente, los datos procesados fueron entregados a los administradores de la Infraestructura de Datos Espaciales de la provincia de Villa Clara (Geocuba), donde se logró publicar información de enfermedades y los síntomas que padecen los autopesquisados. La figura 3 muestra por cada área de salud las cantidades de autopesquisados, que dijeron padecer de enfermedades pulmonares. De igual manera se puede consultar en la página <https://idevida.geocuba.cu/vc>, por área de salud las cantidades de personas que dijeron padecer de fiebre alta,

insuficiencia cardíaca, asma, cáncer, dolor en las extremidades, diarreas, diabetes, tos, dolor de cabeza, debilidad y cansancio, falta de aire, hipertensión, dolor de garganta y enfermedad coronaria. Todo esto constituye una poderosa herramienta para que se tomen mejores decisiones por parte de las autoridades de salud.

Los *scripts* creados permiten agilizar el proceso de generación de mapas, debido a que no es necesario hacer uso de herramientas que requieren interacción con el usuario, como es el caso de *Excel* o algún SIG. También posibilitan el procesamiento de un alto volumen de datos de manera eficiente y eficaz. Una de las principales ventajas de estos *scripts* es que permiten generar más de 400 mapas en cuestión de minutos, esto lo hace muy superior a los SIG para estos casos.

Si bien los *scripts* desarrollados son específicos para esta problemática en particular, pueden generalizarse y adaptarse fácilmente para otros tipos de problemas con diferentes datos que se puedan visualizar sobre otras regiones.

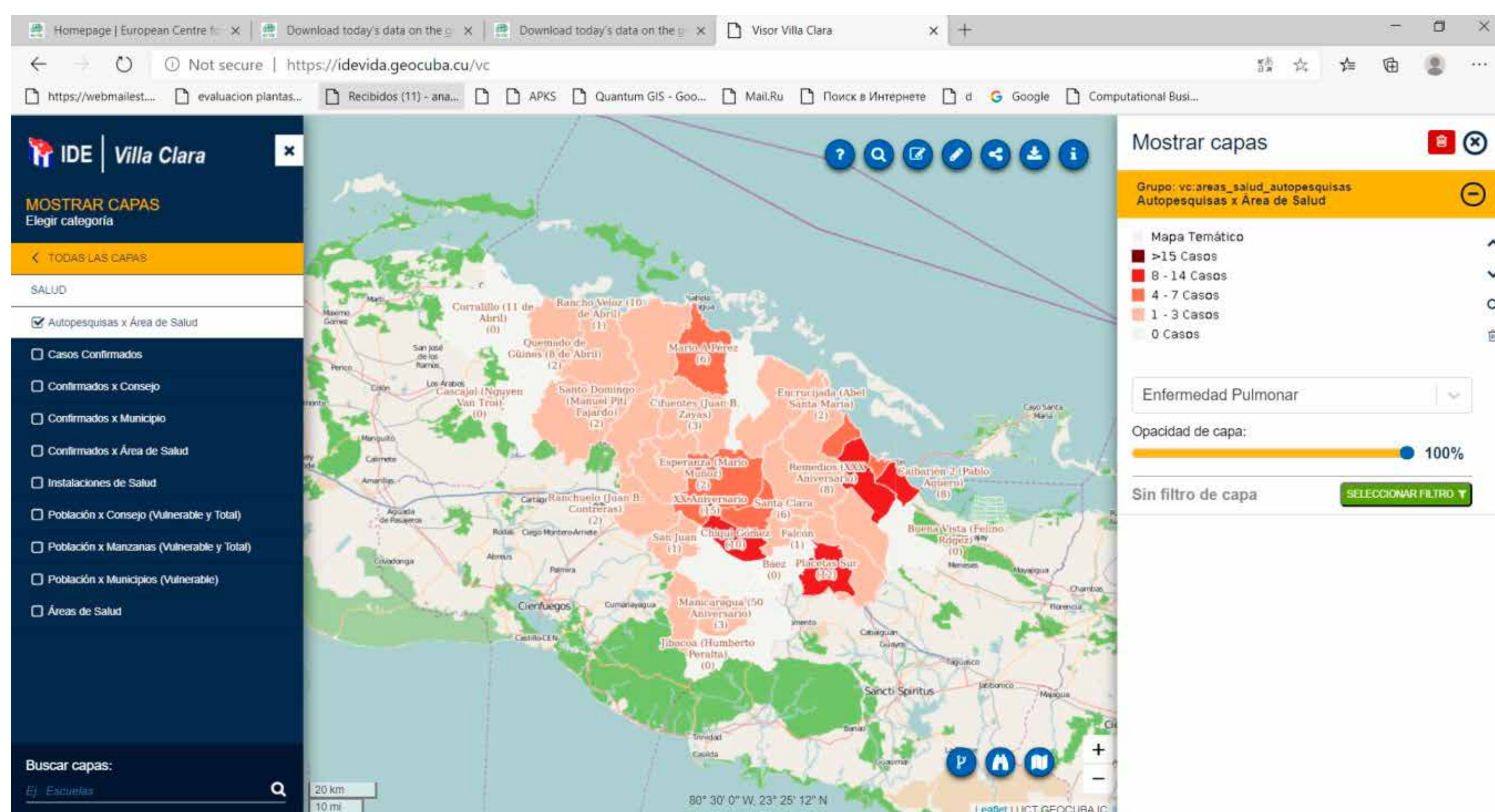


Figura 3. Mapa de coropletas de casos autopesquisados con enfermedad pulmonar.

Los mapas generados pueden ser analizados mediante imágenes independientes, que presentan el escenario de la enfermedad COVID-19 (casos positivos, fallecidos, datos de la autopesquisa) por día y región, o mediante animaciones, que muestran la evolución de los casos a lo largo del tiempo. De esta forma se puede arribar a conclusiones sobre la distribución temporal y espacial de la epidemia en Cuba y el mundo.

La visualización de una o varias regiones al mismo tiempo mediante animaciones permite comparar el desarrollo de la enfermedad en poblaciones cercanas y generar hipótesis sobre la posible relación entre ellas. Así, las autoridades competentes pueden tomar medidas específicas de contención, como, por ejemplo, establecer puntos de control entre regiones o limitar los movimientos de personas de una región específica a otra.

CONCLUSIONES

Se desarrollaron varios *scripts* en R, que permiten procesar y visualizar datos de la pandemia COVID-19 en Cuba y en el mundo, de forma tal que se facilita la realización de análisis espaciales y temporales sobre la evolución de la enfermedad por parte de especialistas e investigadores.

Estos *scripts* se pueden generalizar para ser utilizados en otras áreas de investigación; no están vinculados exclusivamente con la medicina. Fueron diseñados y programados lo más general posible, de manera que se pueden aplicar también a la demografía, sociología, meteorología y muchas otras áreas. También se pueden integrar de forma interactiva con otras herramientas de análisis, como es el caso de los Sistemas de Información Geográfica.

Se generaron más de 400 mapas de coropletas y de símbolos graduados de Cuba y el mundo con datos de la COVID-19, así como animaciones que permiten estudiar la evolución de la enfermedad en el tiempo.

Los *scripts* programados para generar los mapas son resultados importantes de este trabajo, que demuestran la viabilidad del uso del lenguaje R para el procesamiento de datos y la generación de mapas sobre COVID-19.

REFERENCIAS

- Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis* (Vol. 413). Longman Scientific & Technical Essex.
- Barroso, A. R., & Puebla, J. G. (1997). *Los sistemas de información geográficos: origen y perspectivas* (Vol. 7, p. 106). Madrid.
- Bellosta, C. J. G. (2018). R para profesionales de los datos: una introducción.
- Bivand, R. (2013). Package “ classInt ”, 1–12.
- Bivand, R., Lewin-Koh, N., Pebesma, E., Archer, E., Baddeley, A., Bearman, N., ... Rogerbivandnhno, M. R. B. (2016). Package “ maptools ”, 81.
- Bravo, J. D. (2000). Breve Introducción a la Cartografía y a los Sistemas de Información Geográfica (SIG)., (June), 30.
- Díaz-Castrillón, F. J., & Toro-Montoya, A. I. (2020). SARS-CoV-2 / COVID-19 : el virus , la enfermedad y la pandemia, 2019, 183–205.
- Durango Vanegas, C. E. (2013). Caracterización de datos espacio temporales en sistemas de información geográfica. *Instituto de Sistemas y Ciencias de la Decisión*.
- Harries, K. D. (1995). *Mapping crime: Principle and practice*. US Department of Justice, Office of Justice Programs, National Institute of~....
- Hernández-Leal, E. J., Duque-Méndez, N. D., & Moreno-Cadavid, J. (2017). Big Data: an exploration of research, technologies and application cases. *TecnoLógicas*, 20(39), 17–24.
- Hijmans, R. J. (2013). Introduction to the ‘ raster ’ package (version 2 . 1-49), (2008), 1–27.
- Keitt, T., Rowlingson, B., & Bivand, R. (2013). Package “ rgdal ”, 01, 46.

- Lamigueiro, Ó. P. (2014). *Displaying time series, spatial, and space-time data with R*. CRC Press.
- Mas, J.-F. (2018). *Análisis espacial con R : Usa R como un Sistema de Información Geográfica* (p. 151).
- Messy, T. (2019). Package “tidyr.”
- Murphy, L. D. (1995). Geographic information systems: are they decision support systems? In *Twenty-Eighth Annual Hawaii International Conference on System Sciences* (pp. 131–141). Hawaii.
- Neuwirth, E. (2013). Package “RColorBrewer”, 1–5.
- Olaya, V. (2014). *Sistemas de Información Geográfica* (p. 798).
- Paradis, E., & Ahumada, J. A. (2003). R para Principiantes.
- Richards, T. B., Berkowitz, Z., Thomas, C. C., Foster, S. L., Gardner, A., King, J. B., ... Royalty, J. (2010). HERRAMIENTAS & TÉCNICAS El diseño de mapas de coropletas para representar la incidencia del cáncer.
- Salmerón Gómez, R., & others. (2008). Análisis estadístico de datos espacio-temporales mediante modelos funcionales de series temporales.
- Santana, J. S., & Farfán, E. (2014). *El Arte de Programar en R: Un lenguaje para la estadística* (p. 197).
- Sherman, L. W., Gartin, P. R., & Burger, M. E. (1989). Hot spots of predatory crime : routine activities and the criminology of place, (September 2014), 56.
- Stadler, M. M. (2002). *Topología General* (p. 200). Managua. Retrieved from <http://www.ehu.es>
- Taylor, R. B., & Brower, S. (1984). Block Crime and Fear : Defensible Space , Local Social Ties , and Territorial Functioning, (May), 331. doi:10.1177/0022427884021004003
- Vázquez Rodríguez, R. (2015). *Nuevos métodos para el procesamiento y análisis de información geográfica*. Universidad de Granada.
- Wickham, H. (2020). Package “plyr.”
- Wickham, H., & Chang, W. (2015). Package “ggplot2.”

Copyright © 2020 Reinoso-Lorente, A., Vázquez-Rodríguez, R., Pérez-Risquet.



Este obra está bajo una licencia de Creative Commons Reconocimiento 4.0 Internacional.